

PREVISÃO HIDROLÓGICA PROBABILÍSTICA DERIVADA DA MODELAGEM DOS ERROS E DA APROXIMAÇÃO META-GAUSSIANA

Anne Caroline Negrão ^{1*} & Eduardo Alvim Leite ² & Ângelo Breda ³

Resumo – Por considerar as incertezas das previsões, modelos probabilísticos são mais fiéis à aleatoriedade dos fenômenos hidrológicos e agregam valor na tomada de decisão. Buscando adicionar informações estocásticas aos resultados determinísticos do modelo utilizado pelo Sistema de Previsão Hidrológica do Iguaçu (SISPSHI), foi desenvolvido um modelo probabilístico derivado dos erros das previsões, utilizando-se de uma aproximação meta-Gaussiana. Para isso, utilizou-se a Transformação do Quantil Normal (*Normal Quantil Transformation* - NQT) e as estatísticas foram tratadas no domínio Gaussiano. A forma de apresentação probabilística dos resultados agregou valor as previsões e permitiu uma melhor compreensão das incertezas envolvidas.

Palavras-Chave – modelos hidrológicos, previsão probabilística, previsão determinística.

PROBABILISTIC HYDROLOGICAL FORECASTING DERIVED FROM MODELING OF ERRORS AND META-GAUSSIAN APPROACH

Abstract – Probabilistic models consider the uncertainty of forecasts, therefore they are more faithful to the randomness of hydrological phenomena and add value in decision making. In order to add stochastic information to the deterministic results of the model used by Hydrologic Prediction System of Iguaçu (SISPSHI), a probabilistic model derived from the errors of forecasts, using a meta-Gaussian approach, was developed. For this, the Normal Quantil Transformation (NQT) was used and the statistical inference were performed in the Gaussian domain. The form of probabilistic presentation of the results added value to forecasts and allowed a better understanding of the uncertainties involved.

Keywords – hydrological model, probabilistic forecast, deterministic forecast.

INTRODUÇÃO

A previsão hidrológica é uma ferramenta bastante utilizada na solução de problemas de grande impacto social, ambiental e econômico. Porém, mesmo assim, as incertezas envolvidas às previsões não são devidamente consideradas. Apesar do formato predominante de previsão ser determinística, o ideal seria que as previsões fossem expressas em termos de probabilidades (KRZYSZTOFOWICZ, 2001).

Tal observação pode ser exemplificada com a operação de reservatórios. Durante o processo de tomada de decisão, um valor exato dado por um modelo determinístico pode não expressar satisfatoriamente seu significado e, mesmo assim, ser considerado ilusoriamente como correto e único. O erro de tal decisão, devido a falta de maiores informações, pode envolver grandes perdas.

Modelos probabilísticos consideram as incertezas das previsões, sendo mais fiéis à aleatoriedade dos fenômenos envolvidos. São aplicáveis na análise de riscos, na quantificação da

¹ Instituto Tecnológico SIMEPAR, annecnegrao@gmail.com

² Instituto Tecnológico SIMEPAR, alvim@simepar.br

³ Instituto Tecnológico SIMEPAR, angelo@simepar.br

incerteza em termos da distribuição de probabilidades, na análise das probabilidades de ocorrência para diferentes cenários possíveis, etc (KRZYSZTOFOWICZ, 2001). Por outro lado, tais modelos exigem um histórico de dados representativos e intenso processamento computacional. Além disso, sua interpretação nem sempre é fácil e a análise de qualidade não é direta (KATZ; MURPHY, 1997).

O Sistema de Previsão Hidrológica do Iguaçu (SISPSHI) está em constante desenvolvimento pelo Instituto Tecnológico SIMEPAR e pela COPEL. Sua estrutura principal utiliza o modelo hidrológico determinístico 3R, *rainfall-runoff-routing*, e faz previsões de vazão para as sub-bacias do rio Iguaçu. Buscando agregar informações estocásticas aos resultados determinísticos do modelo utilizado pelo SISPSHI, foi desenvolvido um modelo probabilístico com base nos erros observados, utilizando-se de uma aproximação meta-Gaussiana. Esta é uma das diversas técnicas apresentadas por Montanari (2011) de se avaliar quantitativamente a incerteza.

Tal abordagem faz as análises e desenvolve os modelos no espaço Gaussiano. Para isso, os erros e suas variáveis explicativas são normalizados a partir do método *Normal Quantil Transformation* (NQT). No domínio Gaussiano assume-se que os erros podem ser representados pela combinação linear das variáveis explicativas, o que facilita muito as análises estatísticas (MONTANARI, 2011).

METODOLOGIA

A Tabela 1 apresenta os dados utilizados neste trabalho e seus intervalos. Foram utilizados dados das estações de Santa Cruz do Timbó (B5), Fluviópolis (B8) e União da Vitória (B9), com intervalos a cada 6 horas, considerando os horários de 4, 10, 16 e 22 horas.

Tabela 1 – Dados e Intervalos Utilizados.

Dados	Data Inicial	Data Final
Vazão Observada em B5, B8 e B9	04/03/1998	31/12/2011
Chuva Observada em B9	04/03/1998	31/12/2011
Vazão Prevista SISPSHI com Chuva Observada para B9	04/03/1998	31/12/2011
Vazão Prevista SISPSHI com Chuva MBAR para B9	01/01/2010	31/12/2011

Para obter os dados de vazão prevista SISPSHI o sistema foi executado utilizando como entrada a chuva observada como se fosse chuva prevista. Desde 2010 o SISPSHI utiliza como entrada de previsão de chuva as previsões realizadas pelo sistema meteorológico MBAR (Modelo Brasileiro de Alta Resolução) - derivado do *High Resolution Regional Model* (HRM), desenvolvido pelo Serviço Meteorológico Alemão - que também foram utilizadas neste trabalho.

O modelo de erro (ME) foi desenvolvido visando prever os erros de previsão do modelo determinístico utilizado pelo SISPSHI, tendo como saída os erros de previsão ($e_{prev} = Q_{obs} - Q_{prev}$) a cada 6 horas até 72 horas (12 preditantes). Entre as possíveis variáveis explicativas do erro de previsão (preditores) foram selecionadas os seguintes grupos: (1) últimas vazões previstas em B9 para 6, 24, 48 e 72 horas; (2) últimos degraus de vazão de 6 horas previstos em B9 para 6, 24, 48 e 72 horas; (3) últimas vazões observadas em B5, B8 e B9; (4) últimos degraus de vazão de 6 horas observados em B5, B8 e B9; (5) precipitação observada acumulada das últimas 6 horas; (6) erros absolutos médios das 4 últimas previsões de 6 e 24 horas; (7) últimos erros de 6 e 24 horas.

O conjunto de preditores (x_i , para $i = 1, 2, \dots, 19$) e preditantes (e_j , para $j = 1, 2, \dots, 20$) foi dividido em intervalos de treinamento, de 04/03/1998 à 31/12/2008, e teste, 01/01/2010 à

31/12/2011. Destaca-se que o intervalo de treinamento utiliza chuva observada como chuva prevista e o intervalo de teste chuva MBAR.

O intervalo de teste foi reservado e o intervalo de treinamento foi separado em preditores de erros positivos (x_i^+) e preditores de erros negativos (x_i^-). Essa separação será justificada posteriormente.

Através do NQT, tendo como base a distribuição de probabilidade das variáveis da amostra de treinamento, os preditores e preditantes foram transformados, obtendo-se os valores normalizados no domínio Gaussiano (Nx_i^+ , Nx_i^-). Nesse domínio, é possível assumir que os erros estão relacionados linearmente com as suas variáveis explicativas, conforme a Equação 1, que considera o caso positivo, mas também é validada para o caso negativo.

$$Ne_i^+(t) = C_0 + C_1 \cdot Nx_1^+(t) + C_2 \cdot Nx_2^+(t) + \dots + C_i Nx_i^+(t) \quad (1)$$

onde C_i são as constantes ajustadas por regressão linear múltipla.

A transformação dos preditores considera a distribuição completa da amostra de preditores de treinamento. Já na transformação dos preditantes, para os erros positivos considera-se a distribuição da amostra de erros positivos de treinamento e para os erros negativos considera-se a distribuição da amostra de erros negativos de treinamento.

Com preditores e preditantes transformados, foram desenvolvidos, para cada horizonte de interesse, dois tipos modelos: os modelos de erros positivos (ME_j^+), que foram calibrados com os preditores de erros positivos, e os modelos de erros negativos (ME_j^-), que foram calibrados com os preditores de erros negativos. Para obter modelos mais parcimoniosos, considerou-se o quão explicativo são as variáveis. Para isso, utilizou-se o Teste t de Student e apenas os preditores cujos $\Pr(> |t|) < 0,05$ foram considerados (nível de confiança de 95%).

Para verificar se o ajuste da Equação 1 foi satisfatório, foi feita a análise dos resíduos ($Nr^+ = Ne_{obs}^+ - Ne_{prev}^+$ e $Nr^- = Ne_{obs}^- - Ne_{prev}^-$) a partir dos gráficos quantil teórico x quantil empírico. Tanto no intervalo de treinamento como de teste, considerou-se que a normalidade dos resíduos é satisfatória. Dessa forma, pode-se definir sua distribuição no domínio Gaussiano e aplicá-la na previsão probabilística.

Para a composição dos dois modelos de erros (positivo e negativo), considerou-se que os resíduos de ME^+ representam a banda superior e os resíduos de ME^- a banda inferior da distribuição. Porém, a proporção encontrada entre e^+ e e^- não é simétrica e um ajuste nos percentis foi necessário. Tendo como base a proporção de erros positivos, o ajuste foi feito pelas Equações 2 e 3.

$$\text{se } p < (1 - \alpha^+) \rightarrow \hat{p} = 1 - \frac{(1 - \alpha^+) - p}{(1 - \alpha^+)} \quad (2)$$

$$\text{se } p \geq (1 - \alpha^+) \rightarrow \hat{p} = \frac{p - (1 - \alpha^+)}{\alpha^+} \quad (3)$$

onde p é o percentil, \hat{p} é o percentil transformado e α^+ é a proporção de erros positivos.

A Equação 2 representa os percentis da banda inferior e utiliza-se a distribuição dos resíduos de ME^- e a Equação 3 representa os percentis da banda superior e utiliza-se a distribuição dos resíduos de ME^+ . Assim, foram selecionados os resíduos correspondentes às transformações dos percentis 0,01; 0,02; ...; 0,98 e 0,99 ($Nr_{0,01}^-$; $Nr_{0,02}^-$; ...; $Nr_{0,98}^+$; $Nr_{0,99}^+$) e armazenados.

Dada as previsões determinísticas de Ne^+ e Ne^- , soma-se à elas os resíduos correspondentes a cada percentil, obtendo-se uma previsão da distribuição de probabilidades. A previsão probabilística é desnormalizada pelo processo inverso do NQT e apresentada no domínio real de vazões.

A separação dos erros positivos e negativos é aconselhada por Montanari e Brath (2004) e Montanari e Grossi (2008), que obtiveram melhores resultados com essa separação. Segundo eles, quando os erros são tratados juntos, após a transformação NQT, a normalidade não é satisfatória e a linearidade não pode ser garantida para se definir a Equação 1.

Lembrando que o modelo probabilístico foi calibrado utilizando-se chuva observada como chuva prevista, sua aplicação para o intervalo de teste, que utiliza chuva MBAR, teve algumas consequências. No intervalo de teste os erros são maiores e encontram-se mais concentrados na extremidade da distribuição. Consequentemente, os resíduos dos ME s sofrem um desvio da normalidade nesses casos extremos. Espera-se que com o tempo e a atualização dos parâmetros dos modelos esse efeito seja amenizado.

Resumindo o processo de previsão probabilística: (1) os preditores são selecionados e normalizados via NQT, considerando a amostra de preditores do intervalo de treinamento; (2) são feitas as previsões determinísticas utilizando ME_j^+ e ME_j^- ; (3) os resíduos correspondentes a cada percentil são adicionados às previsões; (4) as previsões são desnormalizadas, considerando a amostra de erros positivos para as previsões feitas por ME_j^+ e a amostra erros negativos para as previsões feitas por ME_j^- ; (5) soma-se a distribuição de erros prevista à previsão determinística feita pelo SISPSHI, obtendo-se a previsão probabilística da vazão.

RESULTADOS

A versão probabilística do SISPSHI apresentada neste trabalho encontra-se, desde janeiro de 2013, em operação pelo SIMEPAR. O sistema é executado automaticamente e faz previsões de vazão de 6 em 6 horas, até um horizonte de 72 horas, para União da Vitória, Paraná. Os resultados são apresentados e analisados através de hidrogramas de dispersão de vazão, que consideram os limites de vazão para as probabilidades de ocorrência de 50, 75, 90 e 98%, e da distribuição do complemento de probabilidade acumulada de cada horizonte.

As Figuras 1, 2, 3 e 4 são alguns exemplos de previsões obtidas para diferentes situações hidrológicas. A vazão observada é comparada com as previsões probabilísticas e determinísticas. Nos gráficos (b) é possível extrair a probabilidade da vazão Q ser maior que um valor selecionado q no horizonte de tempo ocorrido ($P[Q > q]$), bem como a distribuição da probabilidade climatológica de tal evento ocorrer.

A Figura 1 é um caso de escoamento de base, situação de melhor comportamento, tanto do modelo determinístico com do modelo probabilístico. Em (a) a vazão observada está muito próxima da mediana da previsão probabilística e da vazão prevista pelo SISPSHI, o que representa um resultado satisfatório do modelo determinístico.

A Figura 2 é uma situação de descida da hidrógrafa. Apesar das previsões terem desviado um pouco do observado, a faixa de 50% conseguiu envolver o observado em todos os horizontes representando uma situação comum ao histórico.

A Figura 3 é uma situação de subida da vazão. A previsão probabilística indicou possível subida, porém ela foi mais intensa do que o esperado, chegando a ultrapassar a faixa de 90%, o que caracteriza uma situação incomum ao histórico (probabilidade de ocorrência abaixo de 10%).

A Figura 4 é um pico de vazão. Essa é uma situação extrema, incomum ao histórico, e o modelo probabilístico previu grande variabilidade para a vazão futura. O observado ficou dentro da faixa de 50% e a incerteza da situação pode ser avaliada previamente.

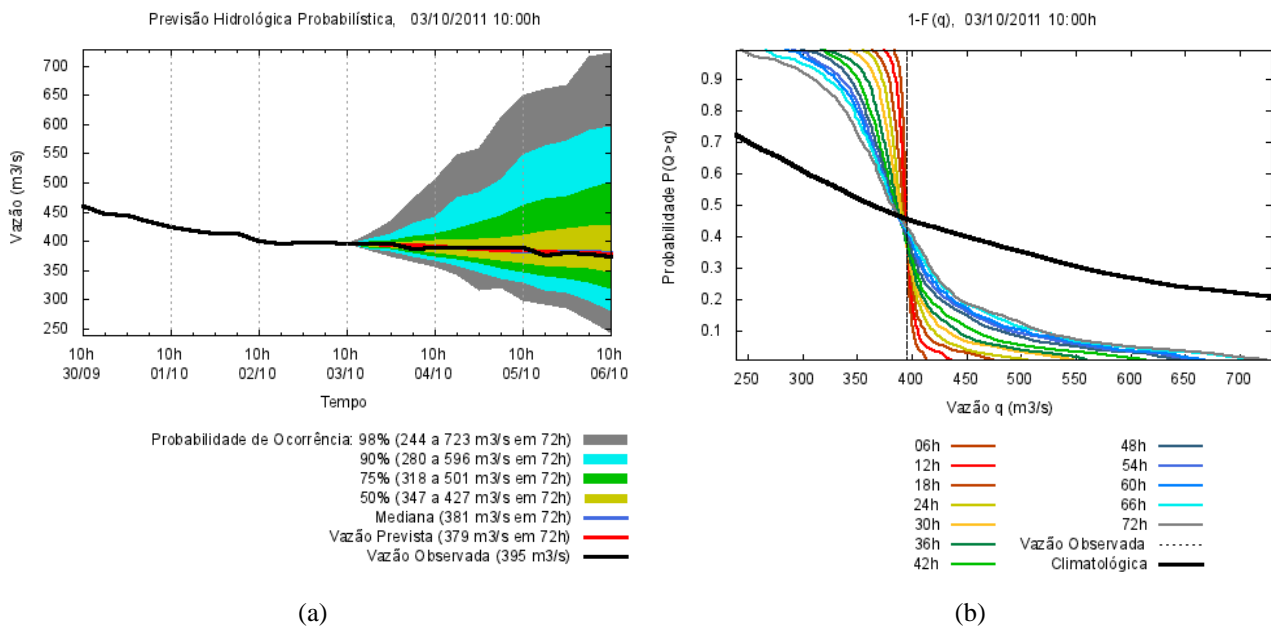


Figura 1 – Previsão hidrológica probabilística feita em 03/10/2011 às 10 horas, (a) hidrograma de dispersão da vazão prevista e (b) distribuição do complemento da probabilidade acumulada da vazão.

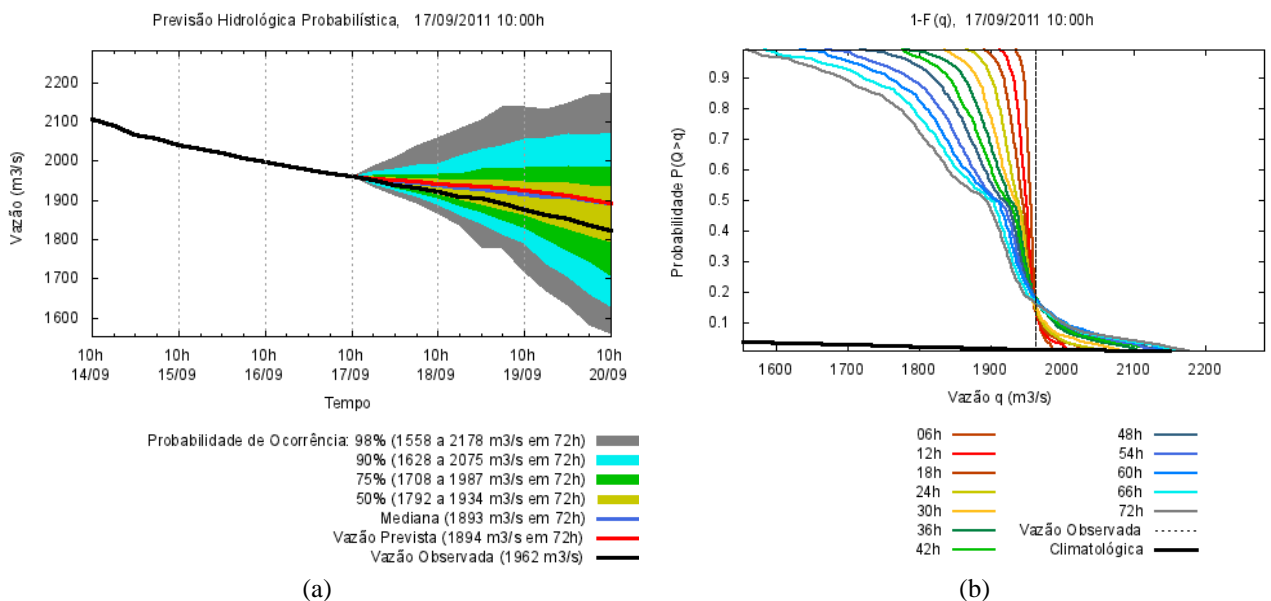


Figura 2 - Previsão hidrológica probabilística feita em 17/09/2011 às 10 horas, (a) hidrograma de dispersão da vazão prevista e (b) distribuição do complemento da probabilidade acumulada da vazão.

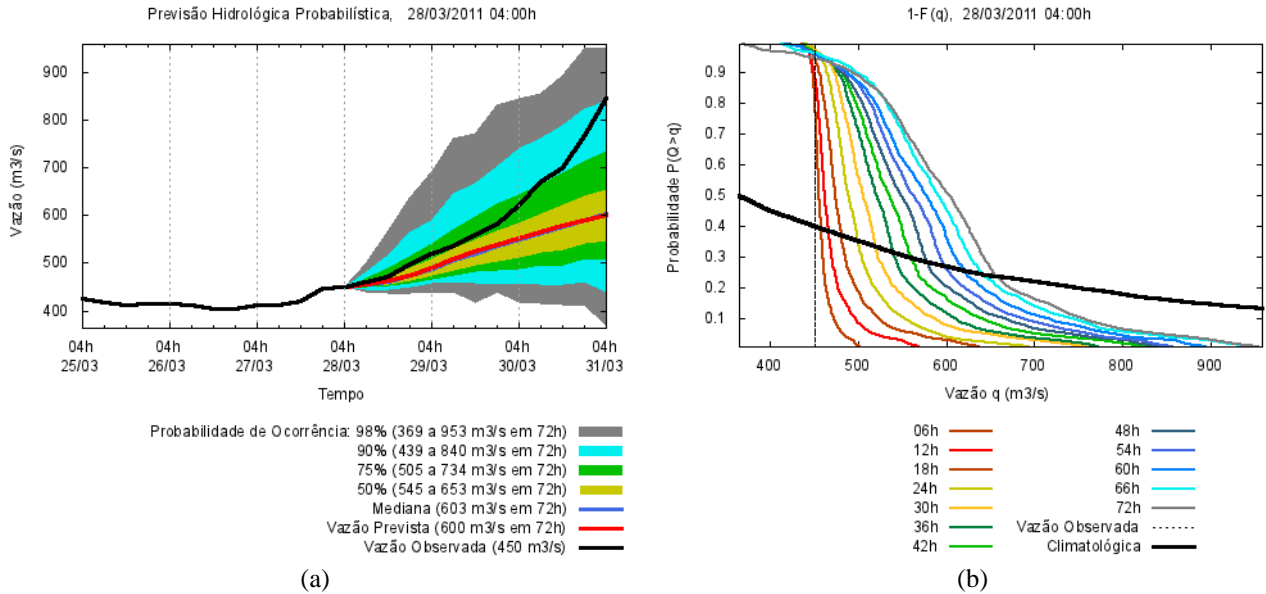


Figura 3 - Previsão hidrológica probabilística feita em 28/03/2011 às 04 horas, (a) hidrograma de dispersão da vazão prevista e (b) distribuição do complemento da probabilidade acumulada da vazão.

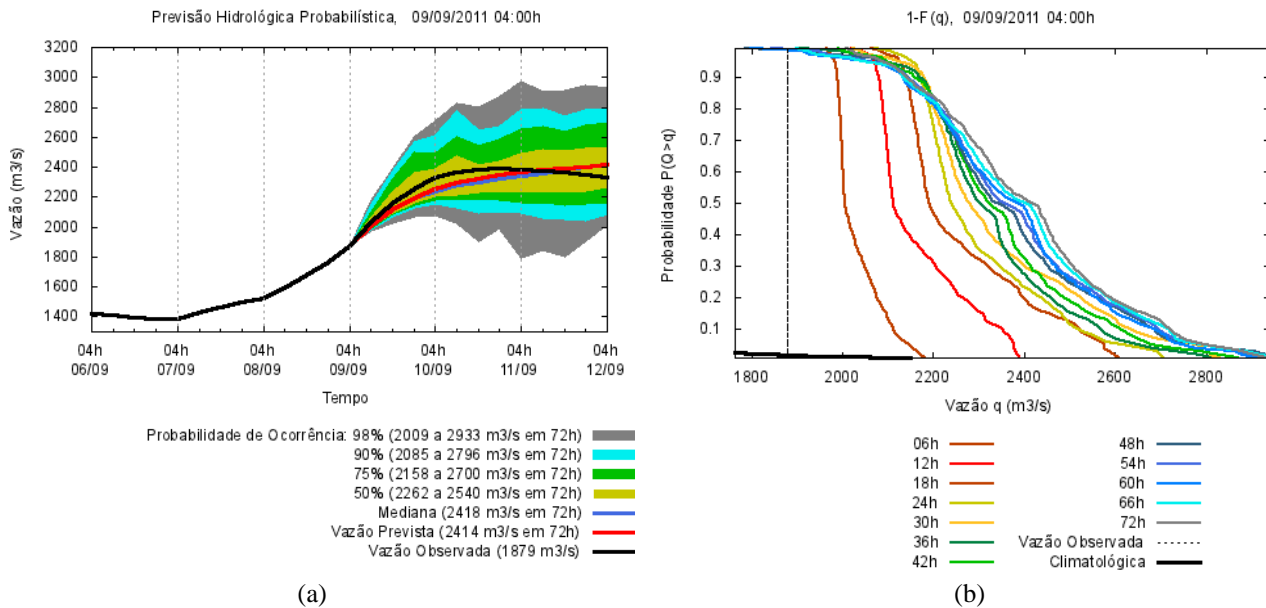


Figura 4 - Previsão hidrológica probabilística feita em 09/09/2011 às 04 horas, (a) hidrograma de dispersão da vazão prevista e (b) distribuição do complemento da probabilidade acumulada da vazão.

CONCLUSÃO

Visando uma descrição mais fiel à realidade estocástica da previsão hidrológica, foi desenvolvido um modelo probabilístico derivado de um modelo determinístico, baseado na estrutura de erros de previsão a partir de uma aproximação meta-Gaussiana. A abordagem probabilística permitiu a explicitação de maiores informações sobre a previsão e disponibilizou as

probabilidades de ocorrência das vazões previstas, o que facilita a tomada de decisão e agrega valor aos resultados.

Os hidrogramas de dispersão de vazão permitem um melhor entendimento das incertezas envolvidas à previsão e favorecem maior discussão durante processo de tomada de decisão. Os gráficos do complemento da distribuição acumulada da probabilidade da vazão permitem derivar a probabilidade de excedência de qualquer vazão selecionada. A partir das curvas de transformação de vazão em nível, todos esses resultados podem expressar também esta variável, fundamental na gestão (controle e alerta de inundações).

Pode-se verificar que o modelo probabilístico acompanha muito bem o modelo determinístico (a mediana é muito parecida com a previsão determinística). Isso já era esperado, pois o SISPSHI tem sido aperfeiçoado no sentido de não apresentar erros sistemáticos. O ganho não foi significativo em relação à qualidade das previsões, mas vale destacar que a abordagem probabilística agrega novas informações a previsão, apresentando tendências variáveis de maior subida ou descida, conforme a situação hidrológica, esse sim foi o verdadeiro ganho.

Uma vantagem da previsão probabilística, que pode ser explorada, é a sua capacidade de evolução e melhoria através de uma atualização dinâmica. Com o passar do tempo novos cenários são armazenados no histórico. Com a entrada de novas informações e a recalibração de parâmetros é possível melhorar a capacidade preditiva do modelo probabilístico.

Além deste modelo apresentado aqui, o SIMEPAR também mantém em operação e desenvolvimento outro modelo hidrológico probabilístico, que tem como base a teoria Baysiana e utiliza redes neurais do tipo SOM (FREIRE, 2009; LEITE, 2011). Uma das dificuldades encontradas na previsão probabilística é fazer a avaliação da qualidade dos modelos. Katz e Murphy (1997) propõem os índices de discriminação e refinamento e Jolliffe e Stephenson (2003) propõem a confiabilidade. Uma metodologia para se quantificar esses índices encontra-se em análise pelos autores, sendo capaz de expressar a qualidade de diferentes modelos probabilístico.

REFERÊNCIAS

- FREIRE, L. S. (2009). Uso de rede neural na obtenção de previsão hidrológica probabilística. *Trabalho Final de Curso de Engenharia Ambiental*, Universidade Federal do Paraná.
- JOLLIFFE, I. T.; STEPHENSON, D. B. (2003). *Forecast verification: a practitioner's guide in atmospheric science*. Chichester, England, John Wiley & Sons.
- KATZ, R. W.; MURPHY, A. H. (1997). *Economic Value of Weather and Climate Forecasts*. Cambridge University Press - USA.
- LEITE, E. A.; BREDÁ, Â.; SABÓIA, J. P. J. (2011). Sistema de previsão probabilística de variáveis hidrológicas e sua aplicação no alerta de inundações. In *Anais do XXI Seminário Nacional de Produção e Transmissão de Energia Elétrica*, Florianópolis, Out. 2011.
- KRZYSZTOFOWICZ, R. (2001). *The case for probabilistic forecasting in hydrology*, Journal of Hydrology, v. 249, n.1-4 (Aug), pp. 2-9.
- MONTANARI, A. (2011). Uncertainty of Hydrological Prediction. In *Treatise on Water Science*. Elsevier. cap. 17. Vol.2, pp.459-478.

MONTANARI, A.; BRATH, A. (2004). A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resources Research*. Vol. 40.

MONTANARI, A.; GROSSI, G. (2008). Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resources Research*. Vol. 44.